Veritism, epistemic risk, and the Swamping Problem

Richard Pettigrew*

October 23, 2018

Abstract

Veritism says that the fundamental source of epistemic value for a doxastic state is the extent to which it represents the world correctly—that is, its fundamental epistemic value is determined entirely by its truth or falsity. The Swamping Problem says that Veritism is incompatible with two pre-theoretic beliefs about epistemic value (Zagzebski, 2003; Kvanvig, 2003):

(I) a true justified belief is more (epistemically) valuable than a true unjustified belief;

(II) a false justified belief is more (epistemically) valuable than a false unjustified belief.

In this paper, I consider the Swamping Problem from the vantage point of decision theory. I note that the central premise in the argument is what Stefánsson & Bradley (2015) call Chance Neutrality in Richard Jeffrey's decision-theoretic framework. And I describe their argument that it should be rejected. Using this insight, I respond to the Swamping Problem on behalf of the veritist.

Veritism, as I will use the term, is a thesis about the purely epistemic value of doxastic states, such as beliefs and credences. In this paper, I will focus only on beliefs, but much of what I say will apply also to credences. Veritism says that the fundamental source of epistemic value for a doxastic state is the extent to which it represents the world correctly—that is, its fundamental epistemic value is determined entirely by its truth or falsity. According to Veritism, any further source of value for a belief is derivative, not fundamental—that is, it

*I am grateful to audiences in Bochum and London, as well as two anonymous referees for this journal, for extremely helpful comments on earlier versions of this paper.

is valuable because of the extent to which it promotes or otherwise serves the fundamental source. The Swamping Problem is an objection to Veritism. It says that Veritism is incompatible with two pre-theoretic beliefs about epistemic value (Zagzebski, 2003; Kvanvig, 2003):

(I) a true justified belief is more (epistemically) valuable than a true unjustified belief;

(II) a false justified belief is more (epistemically) valuable than a false unjustified belief.

Now, Veritism is compatible with the claim that a *true* belief is more (epistemically) valuable than a *false* belief. Indeed, it entails that claim.

And it is also compatible with the claim that a *justified* belief is more valuable than a *unjustified* belief. After all, if we give an account of justification on which a justified belief is more likely to be true than an unjustified belief, as we will below, it follows that, for a veritist, a justified belief has greater expected epistemic value than an unjustified belief; and something with more expected fundamental value is standardly taken to have more derivative value than something with less. For instance, if I find a lottery ticket on the street and I know it comes either from a 10-ticket lottery or from a 100-ticket lottery, both of which pay out the same amount to the holder of the winning ticket, then the situation in which the ticket is from the smaller lottery—and thus boasts a 10% chance of winning—is more valuable than the situation in which it is from the 100-ticket lottery—and thus boasts only a 1% chance of winning—since the former has higher expected monetary value than the latter, and we might suppose that monetary value is the sole fundamental source of value for a lottery ticket.

However, according to the Swamping Problem, Veritism is not compatible with the view that a *true justified* belief is strictly more valuable than a *true unjustified* belief, and it is not compatible with the view that a *false justified* belief is strictly more valuable than a *false unjustified* belief. The argument runs as follows: Suppose I possess a particular item—it might be a lottery ticket; it might be a belief. I take myself to have identified the fundamental source of value for this sort of item; that is, I have identified the variable whose value, once fixed, fixes the fundamental value of the item—if it's a lottery ticket, the fundamental source of its value might be its monetary value; if it's a belief, it might be its truth value, as the veritist contests. Now suppose we are comparing such an item in two different hypothetical situations. In both situations, let's say, the fundamental value is the same—that is, the variable that fixes it takes the same value in both. But the item differs in some other features between these two situations. Then, according to the Swamping Problem, the item cannot be more valuable in one situation than in the other—they must be equally valuable in both. If not, then there must be some other fundamental source of value that accounts for the difference in value—one of the features in which they differ. Their similarity in the feature that determines fundamental value must 'swamp' their difference in any other feature. For instance, suppose again that I find a lottery ticket on the street that I know comes from a 10-ticket lottery or from a 100-ticket lottery, both of which pay out the same amount to the holder of the winning ticket. Then the situation in which the ticket is *the winning ticket* from the smaller lottery is surely exactly as valuable as the situation in which it is *the winning ticket* from the 100-ticket lottery. Although the former had higher expected monetary value than the latter, the fact that it is the winning ticket in both situations, and thus has the same fundamental value in both situations, must 'swamp' any facts about the chance that it would be the winning ticket.

To respond to the Swamping Problem, I want to introduce considerations from decision theory. Although the Swamping Problem is based on considerations of value, it is rarely treated in the decision theory framework. As we will see, considerations from that framework will provide a solution.

The Swamping Problem is usually raised not as an objection to Veritism on its own, but to the conjunction of Veritism with what I will call a *truth-promoting account of justification* (Zagzebski, 2003; Kvanvig, 2003). Examples of such an account include William Alston's indicator reliabilism (Alston, 1988, 2005) and Alvin Goldman's process reliabilism (Goldman, 1979, 2008). On such accounts, a belief in a proposition is justified iff, conditional on some feature of the belief, the proposition believed is likely to be true—that is, its objective probability conditional on that feature is above some threshold $0 < t < 1$. The objective probability to which such accounts allude is a measure of some sort of chance. I won't say more about what sort of chance it is—actual or hypothetical frequentist, perhaps, a propensity, or something else—since my discussion will not depend on the answer. On Alston's account, the feature in question is the ground on which the belief is based. So a belief in a proposition is justified iff, conditional on the agent having the ground on which she in fact bases that belief, the proposition believed is likely to be true. On Goldman's account, the feature is the process by which the belief is formed. So a belief in a proposition is justified iff, conditional on the agent forming the belief using the process she does, the proposition believed is likely to be true.

It is the conjunction of Veritism with one of these accounts of justification that is the usual target of the Swamping Problem, and I will follow that understanding of it here. Thus, I will offer a solution to the Swamping Problem for a veritist with a truth-promoting account of justification. That shows one way in which Veritism can co-exist with (I) and (II).

Let's combine truth-promoting accounts of justification with Veritism. Given Veritism, a true belief is more valuable than a false belief because it has greater fundamental epistemic value. And given the conjunction of Veritism with a truth-promoting account of justification, a justified belief is more valuable than an unjustified belief because it has greater objective expected fundamental epistemic value—that is, greater expected fundamental epistemic value when that is calculated relative to the objective probabilities. But surely a true justified belief is not more valuable than a true unjustified belief, and this for the same reason that a winning ticket from the smaller lottery above is not more valuable than a winning ticket from the larger lottery above—after all, although it has greater objective expected value, that fact is surely 'swamped' by the fact that it is the winning ticket in both cases and so its monetary value is fixed. Or so says the Swamping Problem.

In the presence of a truth-promoting account of justification, the central assumption of the Swamping Problem is this: once the actual fundamental value of an item is fixed, this swamps any facts about the chance that it would have that value. In the decision-theoretic context, H. Orri Stefánsson and Richard Bradley call this assumption Chance Neutrality (Stefánsson & Bradley, 2015). They state it precisely within the framework of Richard Jeffrey's decision theory, and they argue that it is not a requirement of rationality (Jeffrey, 1983). In that framework, the relevant features of your attitudes are represented by a desirability function V and a credence function c, both of which are defined on the same algebra of propositions $\mathcal{F}$. For a proposition A in $\mathcal{F}$, $V(A)$ measures how strongly you desire A, or how greatly you value it, while $c(A)$ measures how strongly you believe A, or your credence in A. The central principle of the decision theory is this:

Desirability Suppose the propositions $X_1, \ldots, X_n$ form a partition. Then

$$V(X) = \sum_{i=1}^{n} c(X_i|X)V(X \,\&\, X_i)$$

That is, roughly, your value for X is your expectation of its value over some partition of X.

Now, suppose the algebra on which V and c are defined includes some propositions that concern the objective probabilities of other propositions in the algebra. Then we suppose throughout that your credence function c obeys David Lewis' Principal Principle (Lewis, 1980):

Principal Principle Suppose the propositions $X_1, \ldots, X_n$ form a partition. And suppose $0 \leq \alpha_1, \ldots, \alpha_n \leq 1$ and $\sum_{i=1}^{n} \alpha_i = 1$. Then

$$c(X_j| \bigwedge_{i=1}^{n} \text{Objective probability of } X_i \text{ is } \alpha_i) = \alpha_j$$

That is, your credence in X_j, conditional on information that gives the objective probability of X_j and other members of a partition to which X_j belongs, should be equal to the objective probability of X_j.

In this framework, Chance Neutrality can be stated as follows:

Chance Neutrality Suppose $X_1, \ldots, X_n$ form a partition. And suppose $0 \leq \alpha_1, \ldots, \alpha_n \leq 1$ and $\sum_{i=1}^{n} \alpha_i = 1$. Then

$$V(X_j \,\&\, \bigwedge_{i=1}^{n} \text{Objective probability of } X_i \text{ is } \alpha_i) = V(X_j)$$

That is, the actual outcome of the chance process that picks between $X_1, \ldots, X_n$ 'swamps' information about the chance process itself in your evaluation, and that evaluation is recorded in your value or desirability function V. A simple consequence of this: if $0 \leq \alpha_1, \alpha'_1 \ldots, \alpha_n, \alpha'_n \leq 1$ and $\sum_{i=1}^{n} \alpha_i = 1$ and $\sum_{i=1}^{n} \alpha'_i = 1$, then

$$V(X_j \;\&\; \textstyle\bigwedge_{i=1}^{n} \text{Objective probability of } X_i \text{ is } \alpha_i) =$$

$$V(X_j \;\&\; \textstyle\bigwedge_{i=1}^{n} \text{Objective probability of } X_i \text{ is } \alpha'_i)$$

That is, X_j coming about as the result of one chance process is exactly as valuable as X_j coming about as the result of some different chance process.

Now consider the particular instance of Chance Neutrality that is cited in the Swamping Problem. Suppose I believe X. I assign greater value to this belief being true and justified than I do to it being true and unjustified; and I assign greater value to it being false and justified than I do to it being false and unjustified. Thus, we have the following restatements of (I) and (II) from above, where Bel_X says that you believe X.

(I*) $V(\text{Bel}_X \;\&\; X \;\&\; \text{Objective probability of } X \text{ is at most } t) <$

$$V(\text{Bel}_X \;\&\; X \;\&\; \text{Objective probability of } X \text{ is greater than } t)$$

(II*) $V(\text{Bel}_X \;\&\; \neg X \;\&\; \text{Objective probability of } X \text{ is at most } t) <$

$$V(\text{Bel}_X \;\&\; \neg X \;\&\; \text{Objective probability of } X \text{ is greater than } t)$$

In both cases, t is the threshold for justification—if the objective probability of X, conditional on whatever feature of the belief figures in the truth-promoting account of justification, exceeds t, then it is justified; if it does not, it is unjustified. Both (I*) and (II*) violate Chance Neutrality.

Thus, in this framework, the argument of the Swamping Problem is just that the following three claims are inconsistent: (I*), a truth-promoting account of justification, and Chance Neutrality. And similarly with (II*) in place of (I*). Thus, if we are to solve the Swamping Problem and retain Veritism alongside a truth-promoting account of justified belief, we must reject Chance Neutrality. To do this, we turn to an argument due to Stefánsson and Bradley (Stefánsson & Bradley, 2015, Section 3). They show that, in the presence of Desirability and the Principal Principle, Chance Neutrality entails a principle called Linearity.[1] Then they claim that Linearity is not a requirement of rationality.

Linearity is the following principle:

Linearity

$$V(\bigwedge_{i=1}^{n} \text{Objective probability of } X_i \text{ is } \alpha_i) = \sum_{i=1}^{n} \alpha_i V(X_i)$$

That is, the value of a lottery is the objective expected value of its outcome. Now, as is well known, real agents often violate Linearity (Buchak, 2013). The most obvious cases are ones like these: let K_1 be a situation in which you receive £20 for sure; and let K_2 be a situation in which a fair coin is tossed and you receive £100 if the coin lands heads and nothing if it lands tails. Thus:

	Heads	Tails
K_1	£20	£20
K_2	£100	£0

Many people value K_1 more than K_2. That is, $V(K_1) > V(K_2)$. But K_1 has lower expected monetary value than K_2. In such cases, defenders of Linearity often note that your utility need not be linear in money. That is, the difference between the utility you assign to £100 and the utility you assign to nothing need not be five times the difference between the utility you assign to £20 and the utility you assign to nothing. If, instead, the former is less than twice the latter, Linearity would then lead you to assign higher expected utility to K_1 than

[1]We make the following abbreviation: P_X^α is the proposition *The objective probability of X is* α. By Desirability,

$$V(\bigwedge_{i=1}^{n} P_{X_i}^{\alpha_i}) = \sum_{j=1}^{n} c(X_j | \bigwedge_{i=1}^{n} P_{X_i}^{\alpha_i}) V(X_j \ \& \ \bigwedge_{i=1}^{n} P_{X_i}^{\alpha_i})$$

By the Principal Principle,

$$c(X_j | \bigwedge_{i=1}^{n} P_{X_i}^{\alpha_i}) = \alpha_j$$

By Chance Neutrality,

$$V(X_j \ \& \ \bigwedge_{i=1}^{n} P_{X_i}^{\alpha_i}) = V(X_j)$$

Therefore,

$$V(\bigwedge_{i=1}^{n} P_{X_i}^{\alpha_i}) = \sum_{i=1}^{n} \alpha_i V(X_i)$$

which is just Linearity, as required.

to K_2. But such a defence does not seem to cover every case in which we prefer a sure thing to an uncertain gamble. Intuitively, it seems rationally permissible even for someone whose utility is linear in money to prefer the sure thing to the gamble on the grounds that she is risk-averse and the gamble risks leaving you empty-handed while the sure thing does not.

What's more, there are cases in which we cannot ensure conformity with Linearity by making the utility function concave in money. The most famous is given by the *Allais preferences* (Allais, 1953). Suppose there are 100 tickets numbered 1 to 100. One ticket will be drawn and you will be given a prize depending on which option you have chosen from L_1, $\dots$, L_4. The table below shows how the four options pay out.

	Tickets 1-89	Tickets 90-99	Ticket 100
L_1	£1,000,000	£1,000,000	£1,000,000
L_2	£1,000,000	£5,000,000	£0
L_3	£0	£1,000,000	£1,000,000
L_4	£0	£5,000,000	£0

Each ticket has an equal chance of winning. Now, it turns out that many people have preferences recorded in the following desirability function V:

$$V(L_1) > V(L_2) \quad \text{and} \quad V(L_3) < V(L_4)$$

That is, they strictly prefer L_1 to L_2 and L_4 to L_3. When there is an option that guarantees them a high payout (£1m), they prefer that over something with 1% chance of nothing (£0) even if it also provides 10% chance of a much greater payout (£5m). On the other hand, when there is no guarantee of a high payout, they prefer the chance of the much greater payout (£5m), even if there is also a slightly greater chance of nothing (£0). The problem is that there is no way to assign values to V(£0), V(£1m), and V(£5m) so that V satisfies Linearity and also these inequalities.[2]

Stefánsson and Bradley show that, in the presence of Desirability and the Principal Prin-

[2]Suppose, for a reductio, that there is. By Linearity,

$$\begin{aligned} V(L_1) &= 0.89V(\text{£1m}) + 0.1V(\text{£1m}) + 0.01V(\text{£1m}) \\ V(L_2) &= 0.89V(\text{£1m}) + 0.1V(\text{£5m}) + 0.01V(\text{£0m}) \end{aligned}$$

ciple, Chance Neutrality entails Linearity; and they argue that there are rational violations of Linearity (such as the Allais preferences); so they conclude that there are rational violations of Chance Neutrality. That is, it is sometimes rationally permissible to value a winning ticket from a smaller lottery differently from a winning ticket from a larger lottery that has the same prize money. And it is sometimes rationally permissible to value a true belief that was very likely to be true differently from a true belief that was very unlikely to be true. In sum: the central assumption of the Swamping Problem, at least as an objection to a truth-promoting account of justification, is false.

Now, we might be tempted to leave our discussion there. After all, the Swamping Problem says that Veritism is incompatible with (I*) and (II*). Chance Neutrality is the key premise in the argument for that conclusion, and we've seen that it is not a requirement of rationality. But by showing that this premise is false, we don't thereby show that the conclusion is also false. To do that, we must say how to set the following epistemic values:

- $V(\mathrm{Bel}_X \ \& \ X \ \& \ \text{Objective probability of } X \text{ is } \alpha)$
- $V(\mathrm{Bel}_X \ \& \ \neg X \ \& \ \text{Objective probability of } X \text{ is } \alpha)$

We will show that it is possible to do this in a way that is in keeping with the reasons for which we abandoned Chance Neutrality, as well as being compatible with Veritism and having the consequences we would like it to have, such as (I*) and (II*). But the method we use will only be one among many possible methods. After all, our goal here is just to show that Veritism and truth-promoting accounts of justification are *compatible* with (I*) and (II*)—so we need only describe one way that makes them all true; it need not be the only legitimate way.

Then, since $V(L_1) > V(L_2)$, we have:

$$0.1V(\text{£1m}) + 0.01V(\text{£1m}) > 0.1V(\text{£5m}) + 0.01V(\text{£0m})$$

But also by Linearity,

$$\begin{aligned} V(L_3) &= 0.89V(\text{£0m}) + 0.1V(\text{£1m}) + 0.01V(\text{£1m}) \\ V(L_4) &= 0.89V(\text{£0m}) + 0.1V(\text{£5m}) + 0.01V(\text{£0m}) \end{aligned}$$

Then, since $V(L_3) < V(L_4)$, we have:

$$0.1V(\text{£1m}) + 0.01V(\text{£1m}) < 0.1V(\text{£5m}) + 0.01V(\text{£0m})$$

And this gives a contradiction.

To do this, let's first return to our example above to see how we might set the values of $V(£20 \ \& \ K_1)$, $V(£100 \ \& \ K_2)$, and $V(£0 \ \& \ K_2)$. This will give us a hint as to how to set the epistemic values we want. Recall, in K_1, you receive £20 with probability 100%; in K_2, you receive £100 with probability 50% and you receive nothing with probability 50%. If we wish to preserve the preference $V(K_1) > V(K_2)$, without appealing to a non-linear utility function, it is most natural to look to a risk-sensitive decision theory, such as Lara Buchak's risk-weighted expected utility theory (Quiggin, 1982, 1993; Buchak, 2013). Also, since we abandoned Chance Neutrality because it entails Linearity and thus cannot accommodate the Allais preferences and other risk-sensitive preferences, and since we seek an account of the epistemic value of justified true and justified false beliefs that is in line with our reasons for abandoning Chance Neutrality, it makes sense to look to a risk-sensitive decision theory. But again, as noted above, there are other ways to save (I*) and (II*) for the Veritist without looking to such a theory.

Buchak states her theory in Savage's framework for decision theory, not in Jeffrey's (Savage, 1954). She responds to the Allais preferences by retaining the analogue of Chance Neutrality and dropping the analogue of Desirability instead. But Pettigrew (2015) shows how to state the very same theory while retaining Desirability and dropping Chance Neutrality instead. The full details of Buchak's theory need not detain us here. I will describe how it works only for cases in which there are just two outcomes: the toss of a coin, for instance, as in the case of K_1 and K_2; or the truth or falsity of a belief, as in our central case. Translated into Jeffrey's framework, Buchak says that, as well as a credence function c and a desirability function V, you also have a risk function r that encodes your attitudes to risk. r takes certain of the probabilities assigned by your credence function and skews them. If r is a risk-averse risk function, it skews the probability of better outcomes by reducing them; if it is a risk-seeking risk function, it increases them; if it is risk-neutral, it keeps them fixed, and standard expected utility is the special case of Buchak's theory in which r is the risk-neutral risk function. Buchak assumes that r is a strictly increasing and continuous function with $r(0) = 0$ and $r(1) = 1$.

To see Buchak's theory in action, consider K_2. The motivation for Buchak's theory, which derives from Quiggin's earlier rank-dependent utility theory, is the following observation: The expected value of a gamble on two outcomes is equal to the value you'll get if the worst outcome comes to pass, plus the extra value over and above that value that you'll get if the best outcome comes to pass *weighted by the probability you'll get that extra value*. The risk-weighted expected value of a gamble on two outcomes is then taken to be the value you'll get if the worst outcome comes to pass, plus the extra value over and above that value that you'll get if the best outcome comes to pass *weighted by the risk-skewed probability you'll get that extra value*. Translated into the framework in which Jeffrey's decision theory is formulated, and not assuming Chance Neutrality, as Buchak does, this becomes:

$$V(K_2) = V(\pounds 0 \,\&\, \overline{\pounds 0}) + r(\frac{1}{2})[V(\pounds 100 \,\&\, \overline{\pounds 100}) - V(\pounds 0 \,\&\, \overline{\pounds 0})]$$

where $\overline{\pounds k}$ is the situation in which you receive $\pounds k$ for certain—that is, with probability 100%. But, as Pettigrew notes, we can rewrite this as an expected utility. First, we let:

- $V(\pounds 0 \,\&\, K_2) = \frac{1-r(\frac{1}{2})}{1-\frac{1}{2}} V(\pounds 0 \,\&\, \overline{\pounds 0})$
- $V(\pounds 100 \,\&\, K_2) = \frac{r(\frac{1}{2})}{\frac{1}{2}} V(\pounds 100 \,\&\, \overline{\pounds 100})$

Second, by Desirability and the Principal Principle, we can then derive:

$$\begin{aligned} V(K_2) &= c(\pounds 100|K_2)V(\pounds 100 \,\&\, K_2) + c(\pounds 0|K_2)V(\pounds 0 \,\&\, K_2) \\ &= \frac{1}{2}V(\pounds 100 \,\&\, K_2) + \frac{1}{2}V(\pounds 0 \,\&\, K_2) \\ &= \frac{1}{2}\frac{r(1) - r(\frac{1}{2})}{1-\frac{1}{2}}V(\pounds 0 \,\&\, \overline{\pounds 0}) + \frac{1}{2}\frac{r(\frac{1}{2})}{\frac{1}{2}}V(\pounds 100 \,\&\, \overline{\pounds 100}) \\ &= (1 - r(\frac{1}{2}))V(\pounds 0 \,\&\, \overline{\pounds 0}) + r(\frac{1}{2})V(\pounds 100 \,\&\, \overline{\pounds 100}) \\ &= V(\pounds 0 \,\&\, \overline{\pounds 0}) + r(\frac{1}{2})[V(\pounds 100 \,\&\, \overline{\pounds 100}) - V(\pounds 0 \,\&\, \overline{\pounds 0})] \end{aligned}$$

which is the value assigned by Buchak's theory.

So, as you can see, the more risk-averse you are, the lower $r(\frac{1}{2})$ is, and so the greater $V(\pounds 0 \,\&\, K_2)$ is and the lesser $V(\pounds 100 \,\&\, K_2)$ is. Also, note that, if you are risk-averse, it is better to receive £0 as a result of a process—such as K_2—that gave you a chance to have something more than it is to receive £0 for sure, while it is worse to receive £100 as a result of a process—such as K_2—that gave you a chance to have something less than it is to receive £100 for sure. Clearly, in this notation, $V(K_1) = V(\text{£}20 \,\&\, \overline{\text{£}20})$.

Thus, for instance, if the value you assign to receiving a sum of money for sure is linear in the amount of money, so that $V(\text{£}k \,\&\, \overline{\text{£}k}) = k$, then $V(K_2) = 100r(\frac{1}{2})$ and $V(K_1) = 20$. So, if $r(\frac{1}{2}) < \frac{1}{5}$, then $V(K_2) > V(K_1)$. This will hold, for instance, if $r(x) = x^3$, which is a risk-averse risk function. If you are risk-neutral, on the other hand, $r(\frac{1}{2}) = \frac{1}{2}$, and thus, $V(K_2) < V(K_1)$.

Let's now apply this to the case of beliefs. Let's suppose that you have a belief in proposition X. Then, according to Pettigrew's redescription of Buchak's theory, we have:

- $V(\text{Bel}_X \,\&\, X \,\&\, P^{\alpha}_X) = \frac{r(\alpha)}{\alpha} V(\text{Bel}_X \,\&\, X \,\&\, P^1_X)$
- $V(\text{Bel}_X \,\&\, \neg X \,\&\, P^{\alpha}_X) = \frac{1-r(\alpha)}{1-\alpha} V(\text{Bel}_X \,\&\, \neg X \,\&\, P^0_X)$

where P^{α}_X means *the objective probability of* X *is* α. So $V(\text{Bel}_X \,\&\, X \,\&\, P^1_X)$ is the utility of being right for certain and $V(\text{Bel}_X \,\&\, \neg X \,\&\, P^0_X)$ is the utility of being wrong for certain. Let

- $R := V(\text{Bel}_X \,\&\, X \,\&\, P^1_X)$, where '$R$' is for 'right'; and
- $W := V(\text{Bel}_X \,\&\, \neg X \,\&\, P^0_X)$, where '$W$' is for 'wrong'.

We need a risk function r, together with R and W, for which the following hold:

(I*) $V(\text{Bel}_X \,\&\, X \,\&\, P^{\alpha'}_X) = \frac{r(\alpha')}{\alpha'} R < \frac{r(\alpha)}{\alpha} R = V(\text{Bel}_X \,\&\, X \,\&\, P^{\alpha}_X)$, for all $0 < \alpha' < \alpha < 1$.

That is, a true belief with higher probability α of being true is more (epistemically) valuable than a true belief with lower probability α' of being true.

(II*) $V(\text{Bel}_X \,\&\, \neg X \,\&\, P^{\alpha'}_X) = \frac{1-r(\alpha')}{1-\alpha'} W < \frac{1-r(\alpha)}{1-\alpha} W = V(\text{Bel}_X \,\&\, \neg X \,\&\, P^{\alpha}_X)$, for all $0 < \alpha' < \alpha < 1$.

That is, a false belief with higher probability α of being true is more (epistemically) valuable than a false belief with lower probability α' of being true.

(III*) $V(\text{Bel}_X \,\&\, X \,\&\, P^{\alpha}_X) = \frac{r(\alpha)}{\alpha} R > \frac{1-r(\alpha)}{1-\alpha} W = V(\text{Bel}_X \,\&\, \neg X \,\&\, P^{\alpha}_X)$, for all $0 < \alpha < 1$.

That is, a true belief with probability α of being true is more (epistemically) valuable than a false belief with probability α of being true.

Suppose you set $R > W > 0$, so that it's better to believe a truth that was guaranteed to be true than to believe a falsehood that was guaranteed to be false. Then we can cook up a risk function r for which (I*), (II*), (III*) hold as follows:

$$r(x) = (ax + (1-a))x$$

where $0 < a < \frac{R-W}{R}$.[3] To see how r behaves, see Figure 1.

Note that, together, (I*), (II*), (III*) entail the following: a true belief is more valuable than a false belief; a justified belief is more valuable than an unjustified belief; a true justified belief is more valuable than a true unjustified belief; and a false justified belief is more valuable than a false unjustified belief. Figure 2 illustrates how the value of a true belief and a false belief increase with their justification for a particular choice of R, W, and a—in particular, $R = 3$, $W = 1$, and $a = 0.25$. In this case, a fully justified false belief is less valuable than a fully unjustified true belief—so, in that sense, the accuracy of the belief trumps its justification when calculating its value.

This, I contend, answers the Swamping Problem. Given a truth-promoting account of justification of the sort that will be particularly appealing to the veritist, we can see why a true justified belief is more valuable than a true unjustified belief. Monetary value is the

[3] *Proof.*

- $r(0) = (a \cdot 0 + (1-a)) \cdot 0 = 0$ and $r(1) = (a \cdot 1 + (1-a)) \cdot 1 = 1$;
- r is continuous and strictly increasing on $[0,1]$;
- Note:
$$V(\text{Bel}_X \,\&\, X \,\&\, P_X^{\alpha'}) = \frac{r(x)}{x}R = (ax + (1-a))R.$$
This is increasing, which gives (I*). And it lies in $[R(1-a), R]$, which we'll use below in the proof of (III*).
- Note:
$$V(\text{Bel}_X \,\&\, \neg X \,\&\, P_X^{\alpha'}) = \frac{1-r(x)}{1-x}W = \frac{1-(ax+(1-a))x}{1-x}W = (ax+1)W.$$
This is increasing, which gives (II*). And it lies in $[W, W(1+a)]$, which we'll use below in the proof of (III*).
- If $a < \frac{R-W}{R}$, then $a < \frac{R-W}{W}$, and thus $R(1-a) > W$ and $R > W(1+a)$. Since both functions are linear, this gives (III*).

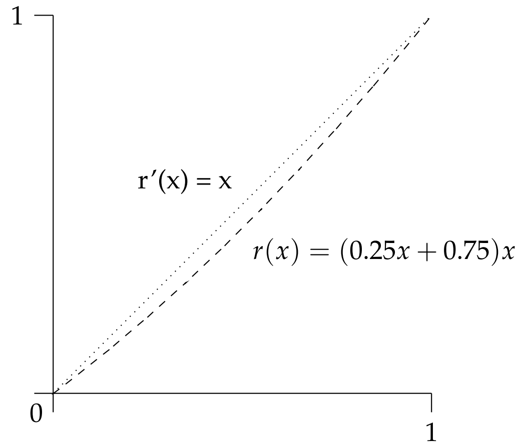

Figure 1: Here, we plot the risk function r (dashed line) that we defined above, picking $a = 0.25$, since this is the value we will go on to use in our example. Alongside it, we plot the neutral risk function r' (dotted line). When applied in Buchak's theory, the neutral risk function returns expected utility theory and Chance Neutrality in Jeffrey's framework. As you can see, r lies always below the neutral risk function, and so it is a uniformly risk-averse function.

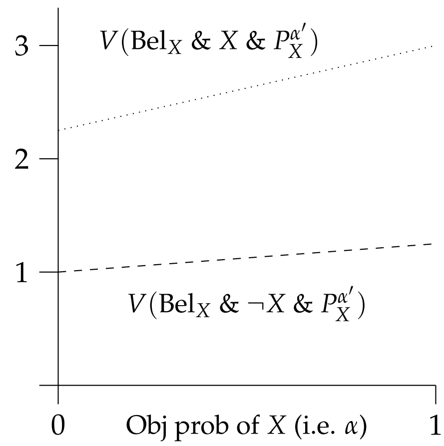

Figure 2

fundamental source of value for a lottery ticket, but we nonetheless assign a lower derivative value to a winning ticket with lower probability of winning than to a winning ticket with higher probability of winning, since we are risk-averse, and valuing them in these ways is the only way to secure our risk-averse preferences in the presence of Desirability and the Principal Principle; similarly, while truth is the fundamental source of value for beliefs, as the veritist contends, we nonetheless assign a lower derivative value to a true belief with lower probability of being true than to a true belief with higher probability of being true. Of course, as we noted above, the particular account of epistemic value here is one among very many that would save (I*), (II*), and (III*). But it demonstrates that it is possible to save these using a decision theory that is designed to accommodate the sort of risk-sensitive preferences that Stefánsson and Bradley show lead to violations of Chance Neutrality.

Before we wrap up, I'd like to consider an objection to this treatment of the Swamping Problem.[4] In Buchak's treatment of her theory, she explicitly keeps the utility function, the risk function, and the credence function separate. For Buchak, attitudes to risk are separate from beliefs and separate from values. This distinguishes her theory from other risk-sensitive theories, which try to do one of two things: they try to incorporate the risk attitudes into the credences—if you're risk-sensitive, then you're portrayed as pessimistic and thus think the worst outcome of a two-way gamble is more likely and the best outcome less likely than your evidence suggests; or they try to build the risk attitudes into the utilities—if you're risk-averse, you value the best outcome of a two-way gamble more when it is the result of a process that makes it more likely. Pettigrew's reformulation of Buchak's theory is a theory of the second sort: he incorporates the risk attitudes into the utilities; he takes the outcomes to specify both means and ends, not just ends; thus, since outcomes are the source of fundamental value in the Savage framework in which Buchak's theory and Pettigrew's reformulation is stated, fundamental value must attach to both means and ends. And of course, we have adopted Pettigrew's reformulation. So it might seem that the same is true of us. But that would be incompatible with Veritism, which says that fundamental value attaches only to the end, namely the truth or falsity of the belief, not the means by which that

[4]Thanks to two anonymous referees for this journal, as well as audiences in Bochum and London for pushing me to answer this concern.

end is achieved, namely, the grounds on which the belief is based for Alston or the process by which it is formed for Goldman.

The first thing to say is that we have not only adopted Pettigrew's reformulation; we have also transferred it over to Jeffrey's framework from the Savage framework in which it and Buchak's original theory are stated. This makes a significant difference. Unlike in Savage's framework, in Jeffrey's framework, utility doesn't just attach to the outcomes; it doesn't just attach to the sources of fundamental value. Rather, in Jeffrey's framework, the desirability function assigns value to *any* proposition about which the agent has an opinion; not just those that determine fundamental value. So, while Buchak's utility function, and the reformulated utility function in Pettigrew's redescription assign value only to outcomes that are specified exactly enough to fix fundamental value—no more and no less—the transposition into Jeffrey's framework does not. For instance, Jeffrey's framework assigns a value to the proposition *Ticket 5*, which says that I have ticket number 5 in a 10-ticket lottery. But that proposition doesn't determine enough to specify fundamental value. Rather, by Desirability,

$$
\begin{aligned}
&V(\textit{Ticket 5}) = \\
&\qquad P(\textit{Ticket 5 wins}|\textit{Ticket 5})V(\textit{Ticket 5} \,\&\, \textit{Ticket 5 wins}) + \\
&\qquad\qquad P(\textit{Ticket 5 loses}|\textit{Ticket 5})V(\textit{Ticket 5} \,\&\, \textit{Ticket 5 loses})
\end{aligned}
$$

where *Ticket 5* & *Ticket 5 wins* and *Ticket 5* & *Ticket 5 loses* both specify enough to detemine fundamental value, which in this case is monetary value. So the fact that V assigns value to propositions that specify the means as well as the ends, and indeed assigns different values to those propositions and to ones that just specify the ends, does not necessarily show that the means must also be a source of fundamental value.

But that doesn't answer the worry completely. The concern might be not that our transposition into the Jeffrey framework of Pettigrew's reformulation of Buchak entails that the desirability function applies to propositions that specify the means as well as ends. Rather, the concern might be that the values it assigns to those propositions don't relate in the correct way to the values it assigns to propositions that just specify the ends. In particular, you might take it simply to be a consequence of saying that something is the only source of fundamental value that, once it is specified, specifying anything further should not change the value. And of course a particular case of this is Chance Neutrality—once the outcome of a chance process is specified, specifying further the process that produced it should not change the value. But the lesson of Stefánsson and Bradley's insight is that this claim is false. When I am risk-averse or risk-seeking, it doesn't hold. It doesn't hold because, being risk-averse or risk-seeking forces me to violate Chance Neutrality. But that is not because a risk-averse or risk-seeking agent is forced to assign *fundamental* value to the means by which the end is achieved. Rather, those means have *derivative* value precisely because they are means to the end, and the end determines the *fundamental* value. The means are only valued in the way they are because they are means to the end, which has fundamental value. What Stefánsson and Bradley's treatment of Chance Neutrality teaches us is that the way we assign derivative value to different ways of promoting fundamental value need not satisfy Chance Neutrality. And indeed it will do so only when you are risk-neutral. When my utility is linear in money, but I prefer a sure £20 to a coin toss between £100 and £0, this is not because I assign fundamental value to a more certain means to my end—it is simply because I am risk averse, and this forces the derivative value of the means to my end to be fixed in a particular way that violates Chance Neutrality.

One interesting upshot of this treatment of the Swamping Problem is that it does not make it *compulsory* to prefer a justified true belief to an unjustified true belief; it only makes it *permissible*. After all, it is not compulsory to be risk-averse, and certainly not compulsory to be risk-averse in one of the ways that ensure (I*), (II*), and (III*) from above. If you are risk-seeking (so that $r(x) > x$, for all x) or even risk neutral (so that $r(x) = x$, for all x), for instance, you will not have that preference.

Thus, I conclude: the Swamping Problem can be answered. It relies on Chance Neutrality, but that is not a requirement of rationality. And once we see why it fails, we can give an account of a desirability function in Richard Jeffrey's decision-theoretic framework that measures epistemic value and that captures (I) and (II), as required.

References

Allais, M. (1953). Le comportement de l'homme rationnel devant le risque: critique des postulats et axiomes de l'école Américaine. *Econometrica*, *21*(4), 503–546.

Alston, W. (1988). An Internalist Externalism. *Synthese*, *74*, 265–83.

Alston, W. (2005). *Beyond "Justification": Dimensions of Epistemic Evaluation*. Ithaca, NY: Cornell University Press.

Buchak, L. (2013). *Risk and Rationality*. Oxford University Press.

Goldman, A. (1979). What is justified belief? In G. S. Pappas (Ed.) *Justification and Knowledge*. Dordrecht: Reidel.

Goldman, A. (2008). Immediate justification and process reliabilism. In Q. Smith (Ed.) *Epistemology: New Essyays*, (pp. 63–82). New York: Oxford University Press.

Jeffrey, R. C. (1983). *The Logic of Decision*. Chicago and London: University of Chicago Press, 2^{nd} ed.

Kvanvig, J. (2003). *The Value of Knowledge and the Pursuit of Understanding*. Cambridge: Cambridge University Press.

Lewis, D. (1980). A Subjectivist's Guide to Objective Chance. In R. C. Jeffrey (Ed.) *Studies in Inductive Logic and Probability*, vol. II. Berkeley: University of California Press.

Pettigrew, R. (2015). Risk, rationality, and expected utility theory. *Canadian Journal of Philosophy*, *45*(5-6), 798–826. Unpublished manuscript.

Quiggin, J. (1982). A Theory of Anticipated Utility. *Journal of Economic Behavior and Organization*, *3*, 323–43.

Quiggin, J. (1993). *Generalized Expected Utility Theory: The Rank-Dependent Model*. Kluwer Academic Publishers.

Savage, L. J. (1954). *The Foundations of Statistics*. John Wiley & Sons.

Stefánsson, H. O., & Bradley, R. (2015). How Valuable Are Chances? *Philosophy of Science*, *82*, 602–625.

Zagzebski, L. (2003). The search for the source of the epistemic good. *Metaphilosophy*, *34*(12-28).